Computational Network Biology

Biostatistics & Medical Informatics 826

Fall 2018

Sushmita Roy

sroy@biostat.wisc.edu

https://compnetbiocourse.discovery.wisc.edu

Sep 6th 2018

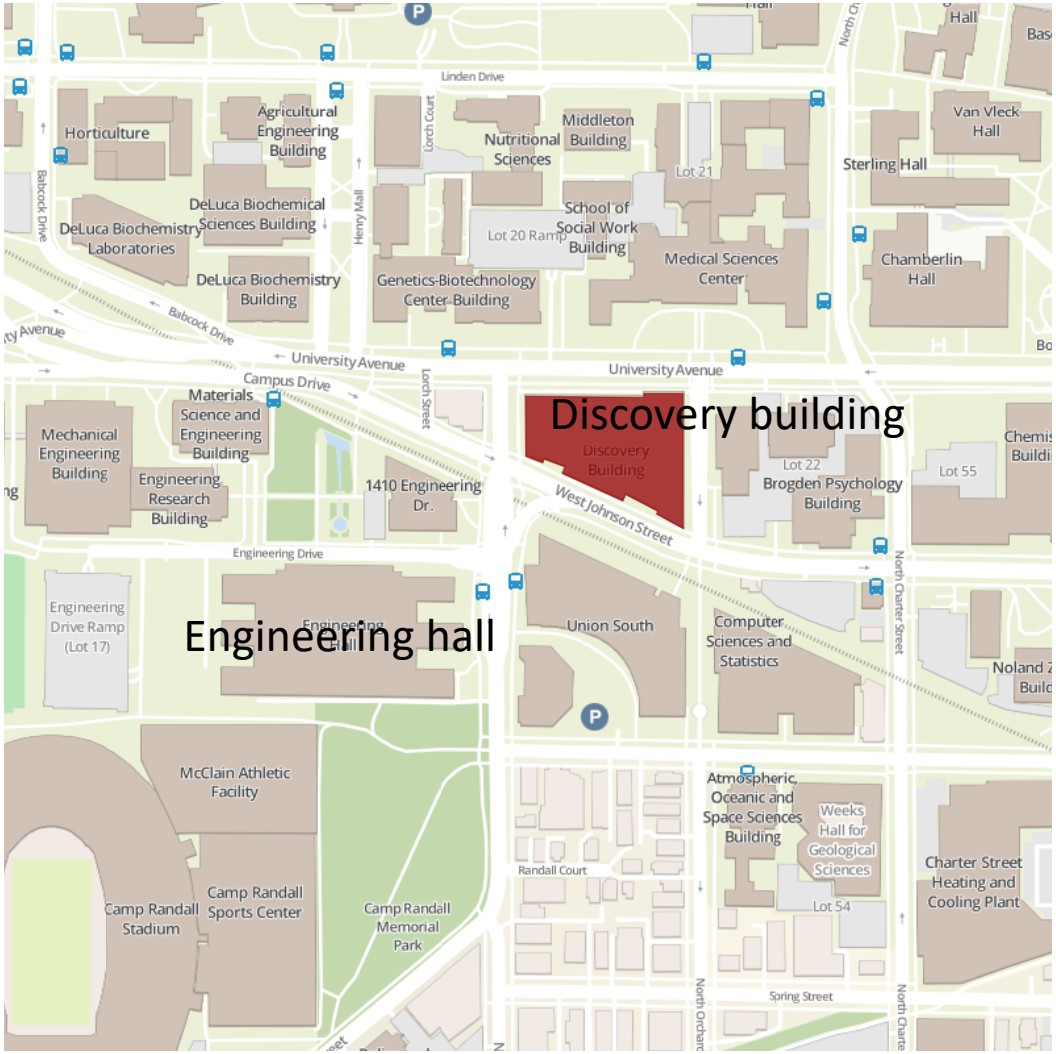
Goals for today

- Administrivia
- Course topics
- Short survey of interests/background

BMI826 Computational Network Biology

- Course home page: **https://compnetbiocourse.discovery.wisc.edu**
- Instructor: Prof. Sushmita Roy
  - sroy@biostat.wisc.edu
- Office: room 3168, Wisconsin Institute for Discovery
  - Your WISC cards will be enabled for upper floor access.
- Office hours:
  - Tuesday/Thursday: 2:30pm-3:30 pm
  - By appointment via email
- Class announcements via piazza
  - Enroll at https://piazza.com/wisc/fall2018/bmi826023

Finding my office WID 3168



Course organization

- Tentative schedule: https://compnetbiocourse.discovery.wisc.edu/schedule-2/
- The material in this course is organized into five major topics
- At the beginning of each topic I will provide an introduction for the topic
- Most of the material is from published papers and review articles
  - We will read and discuss papers from each of these topics
    - Please see syllabus at https://compnetbiocourse.discovery.wisc.edu/syllabus/
  - Readings will be made online on the schedule page.
- Last week or so will be project presentations

Recommended background

- Computer science
  - Introductory course in data structures is good, but not required
- Statistics
  - Good if you've had at least one course, but not required
- Molecular biology
  - Good if you have had some introductory course
  - An interest in learning some basic molecular biology
- Programming background
  - Familiarity with a Linux environment
  - Be able to run programs on data on the command line
  - Be able to write code to do some data analysis and computations

Course grading

- Written critiques: 20%
  - Five written critiques: one for each major topic
- Written and implementation assignments: 30%
  - Three or so
- Project: 45%
  - Proposal 10%
  - Report 20%
  - In class presentation 15%
- In class participation: 5%

Writing a critique

- Critiques are due at the end of each major topic
- Critiques will be a 1-2 page analysis of the papers read in each topic
  - Specific papers to be included in the critique will be mentioned
- The critique should have the following components
  - Overview of the problem area
  - Approaches discussed
  - Strengths and weaknesses
  - Extensions to any of the approaches

Project information

- There are three main components to the project
  - Proposal, In class presentation, Project report
- Project proposal draft (Oct 4th)
- Project proposal final draft (Oct 18th)
- Project presentations in last week of class
- Project report due (Dec 12th)
  - Last day of lecture

Computational resources for this class

- Linux server “mi1.biostat.wisc.edu” available through the BMI department
- Please connect to mi1.biostat.wisc.edu
- Accounts for all registered students in the class have been requested

Goals for today

- Administrivia
- **Course topics**
- Short survey of background and interests

What is Network biology?

- A collection of algorithms and tools to build, interpret, and use graph representations of interacting molecular entities in biological and bio-medical problems (*Adapted from Winterbach et al, 2013, BMC Systems Biology*)
  - ~15 years old
  - The term Network biology was likely coined by Albert-László Barabási & Zoltán N. Oltvai 2004
  - Intersects with computer science, statistics, physics, molecular biology
- Related/overlapping areas
  - Bioinformatics, Systems biology, Complex systems, Biological network analysis, Network science

Why Network biology?

- Cells are complex systems
  - A complex system: many components that interact to determine overall function
  - Networks are natural representations of complex systems
- Provides a framework and important tools for integration, interpretation and discovery
- Many biological applications e.g.
  - Understanding complex biological processes at the molecular level
  - Disease prognosis
  - Interpretation of genetic variation
  - Predictive models of cellular function
  - Gene function prediction and prioritization

Overview of lecture topics

Course material is organized by the biological problem and computational approaches to address the problem

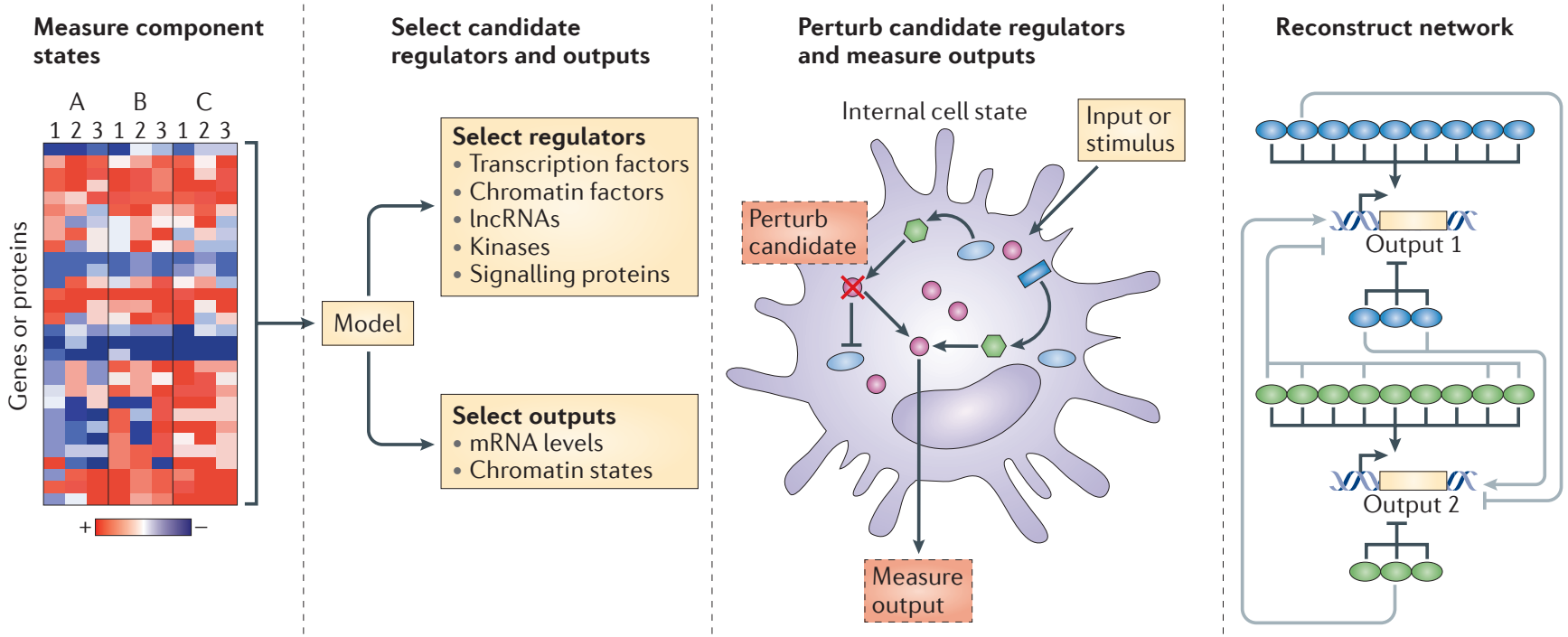
Biological problem

- ❑ Mapping regulatory network structure
- ❑ Dynamics and context specificity of networks
- ❑ Understanding design principles of biological networks
- ❑ Interpretation of sequence variants
- ❑ Identification of important genes
- ❑ Predicting the function of a gene

Computational approaches

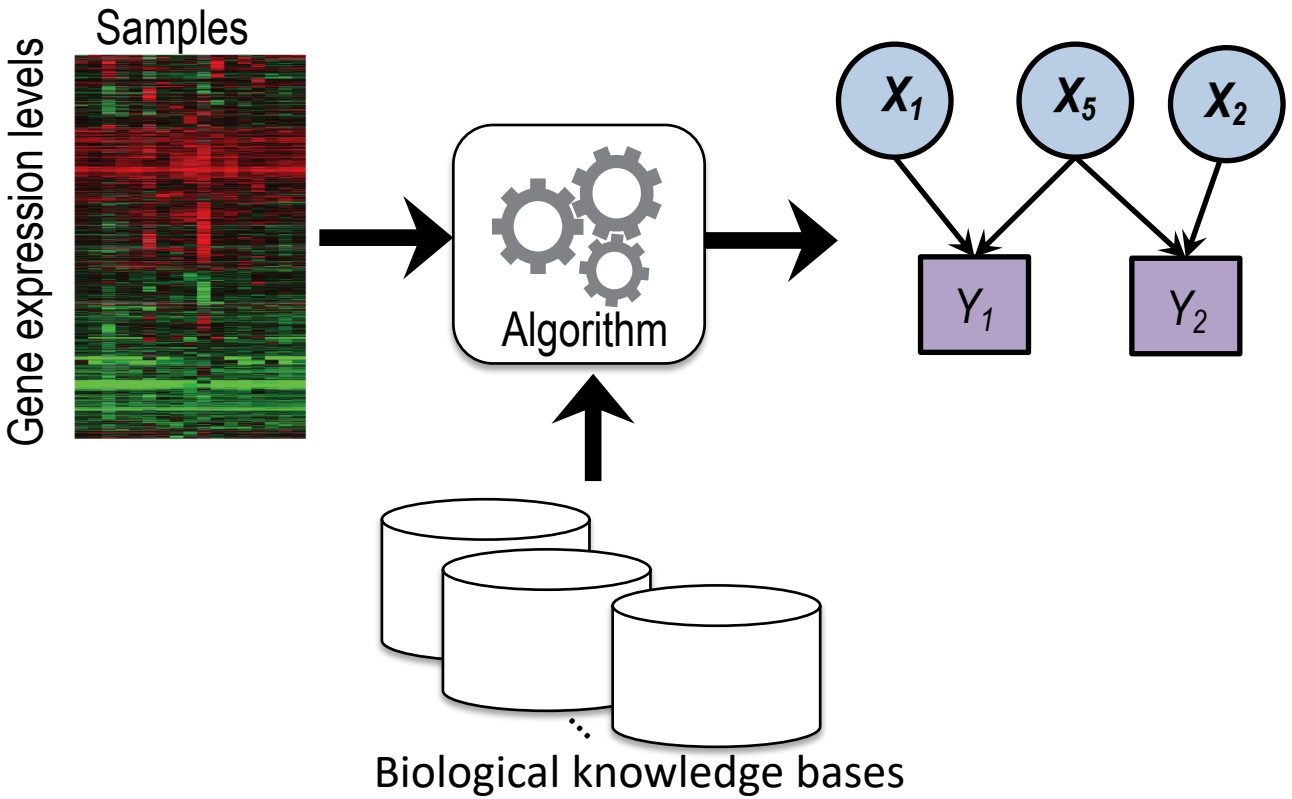
- ❑ Probabilistic graphical models
- ❑ Graph structure learning
- ❑ Multiple network learning
- ❑ Graph clustering
- ❑ Graph alignment
- ❑ Diffusion on graphs

Network inference: How do molecular entities interact within a cell?



Amit et al., Nat. Rev. Immunology, 2011

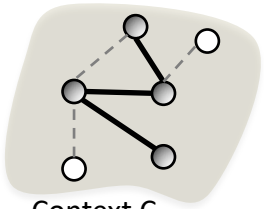
Network inference



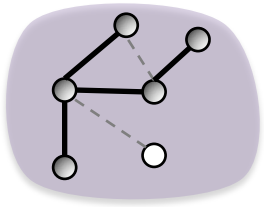
Computational concepts

1. Different types of graphical models for network representation
2. Learning graphical models from data
3. Integrating prior information into models

Network dynamics: How do networks change between different biological contexts?

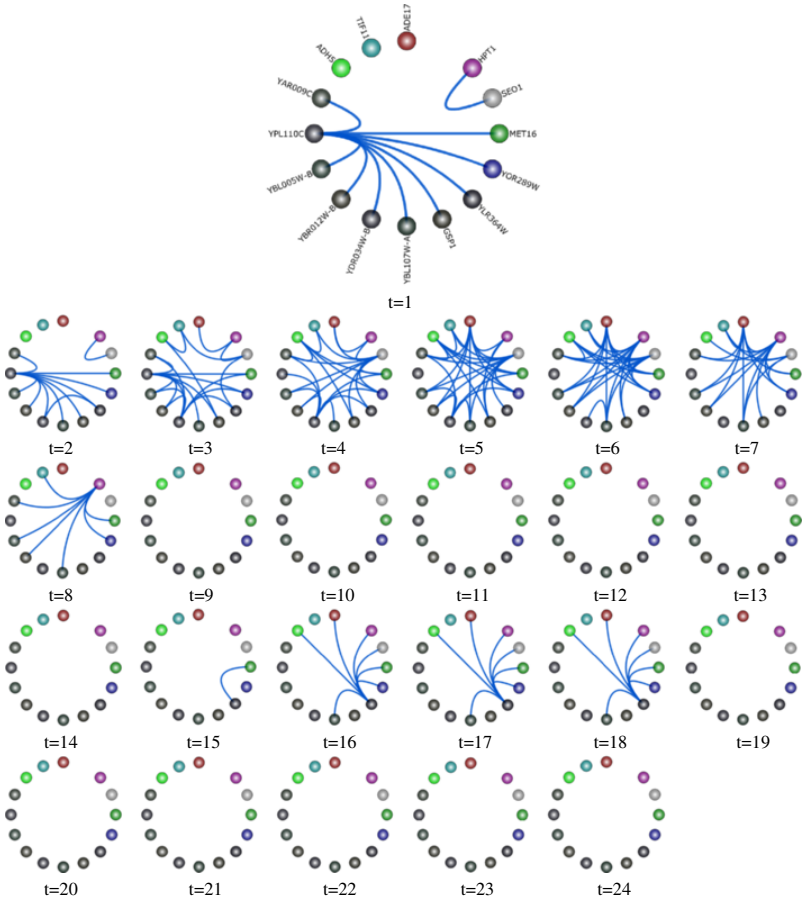


Context C_1



Context C_2

Contexts can be different time points, cell types, disease states, organisms



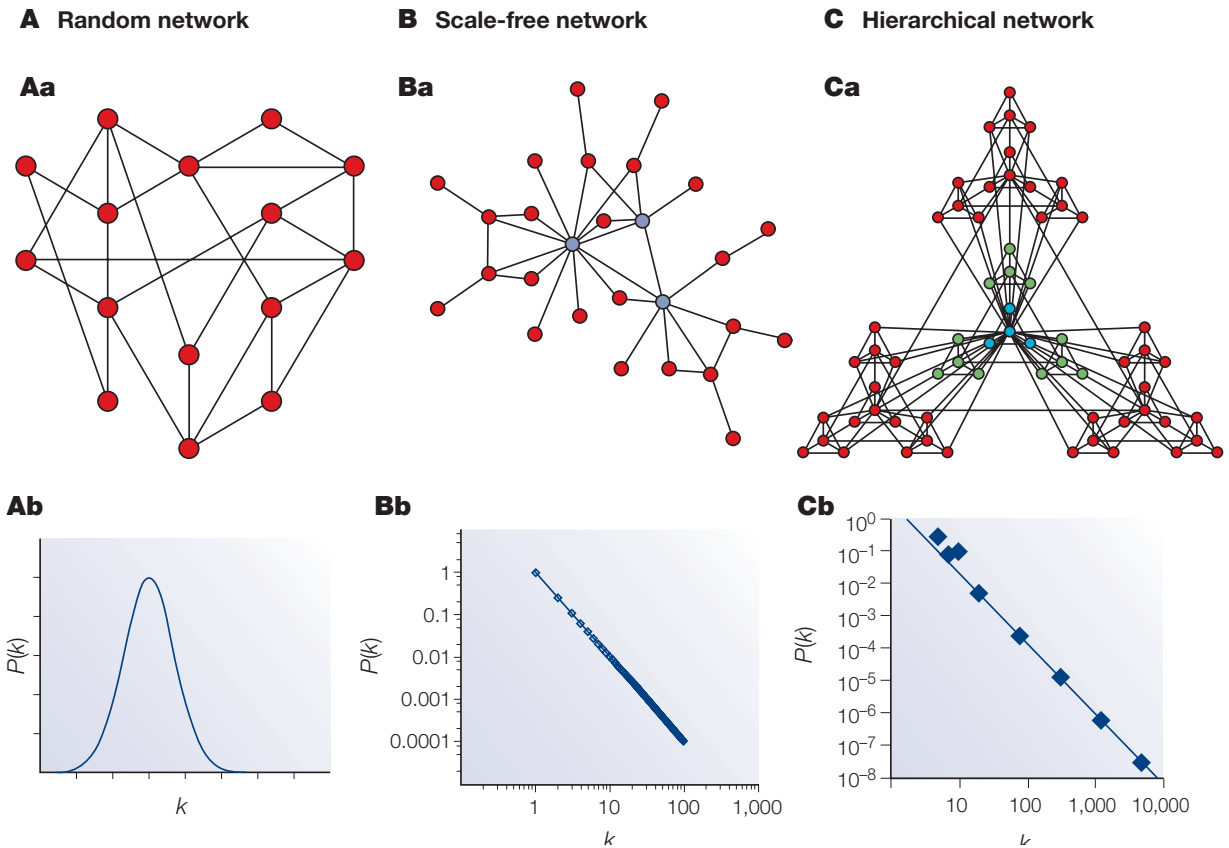
Gene network rewiring during cell cycle

From Curtis et al., BMC Bioinformatics 2012

Computational concepts

1. Multi-task learning
2. Dynamic models for networks

Network topology: How is a network organized?

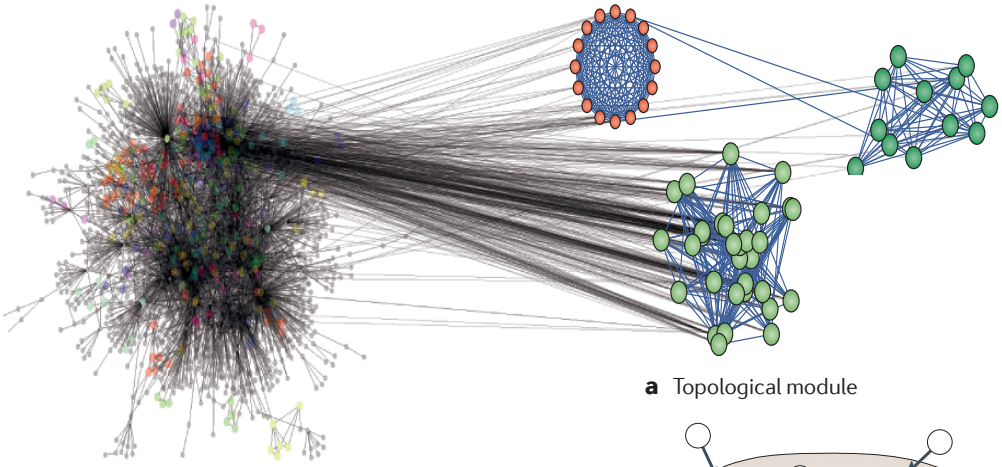


Computational concepts

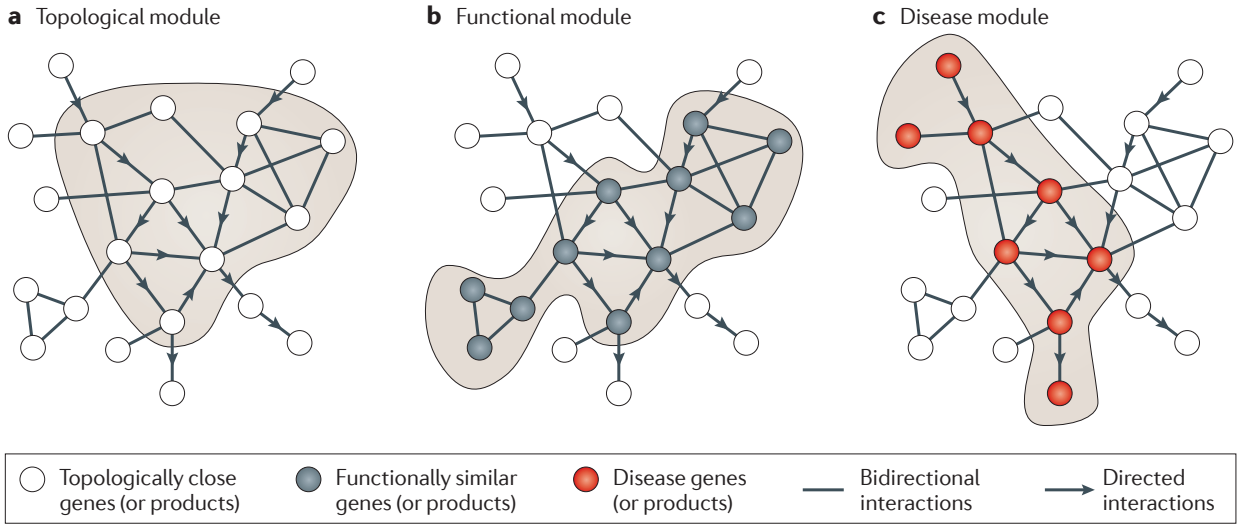
1. Degree distribution
2. Network motifs
3. Centrality measures

Barabasi & Oltvai, Nature Review Genetics 2004

Graph clustering: functional and disease module identification



Mitra et al., Nat Rev Genetics 2013;

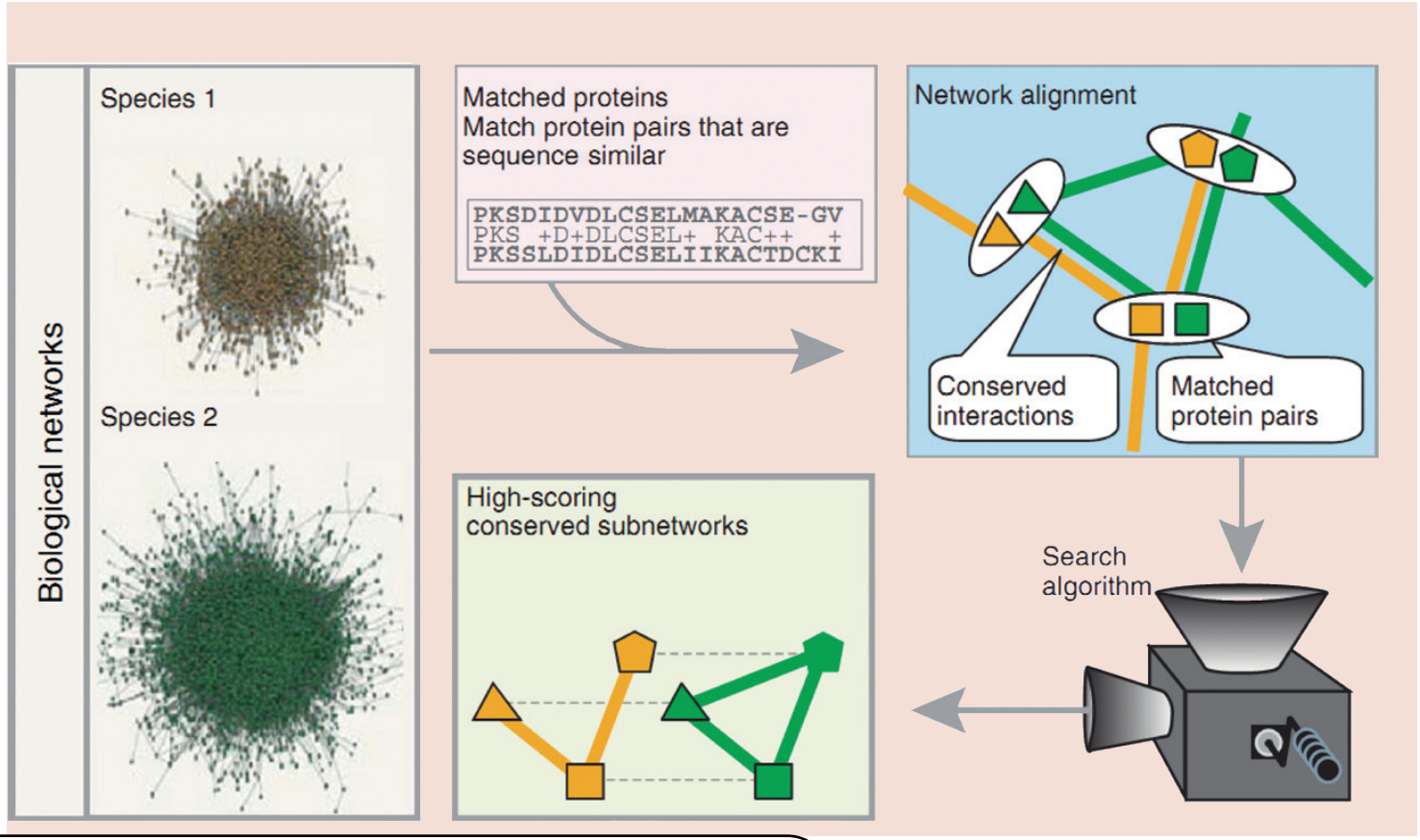


Barabasi et al., Nat Rev Genetics 2011

Computational concepts

1. Graph clustering
2. Modularity measures

Graph alignment: What parts of networks from two species are similar?

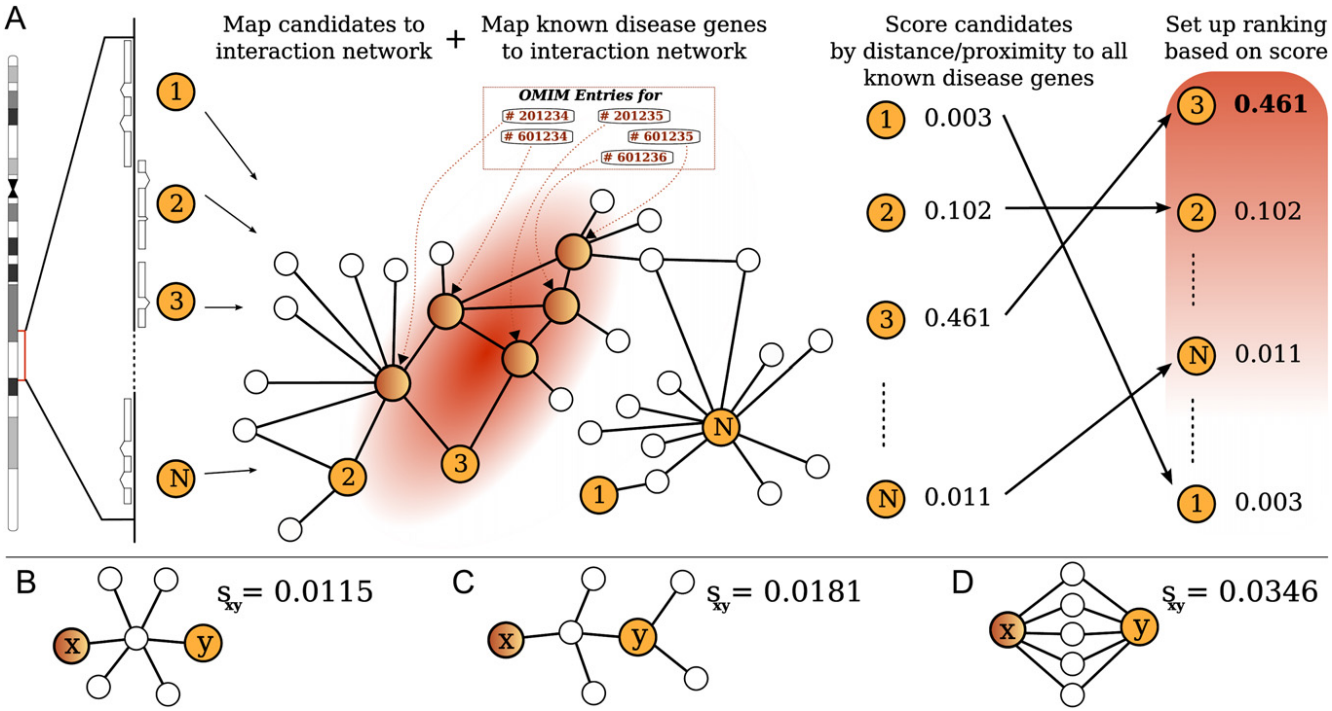


Computational concepts

1. Clustering on graphs
2. Scoring subnetworks and subnetwork search
3. Matrix factorization

Kelley et al PNAS 2003

Graph diffusion: Which genes are most important?

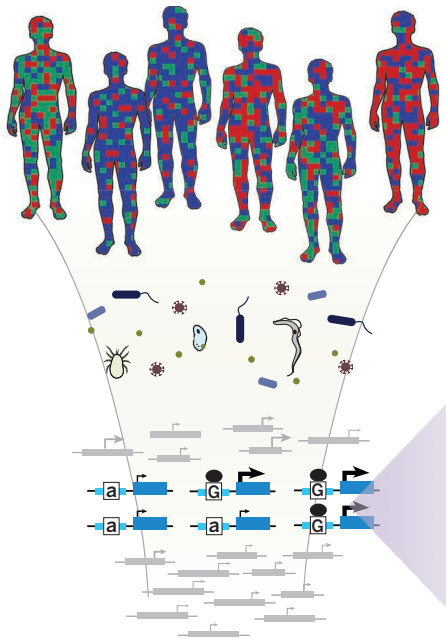


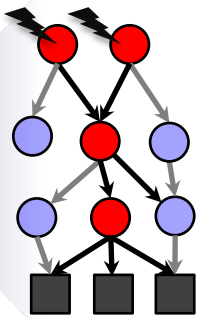
Computational concepts

1. Random walks on graphs
2. Graph diffusion kernels
3. Random walks on graphs

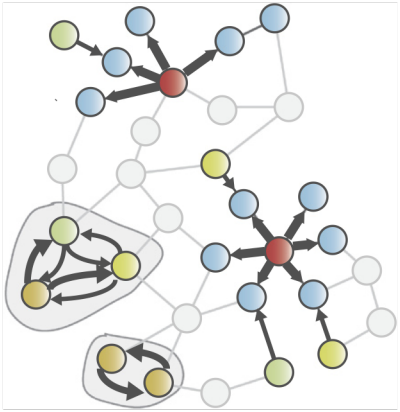
Koehler et al., AJHG 2008

Graph diffusion: Characterizing genetic variation and impact on complex phenotypes

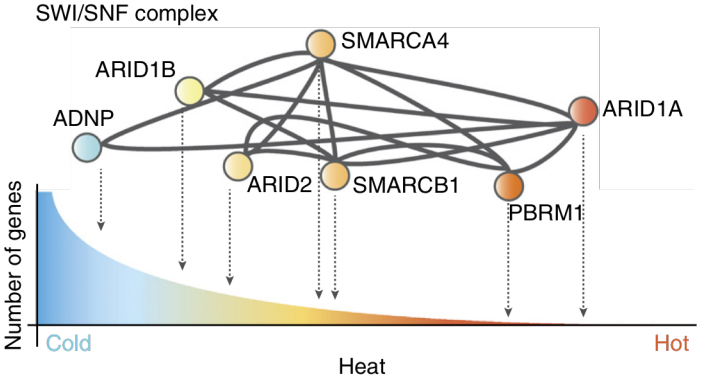




HotNet2 algorithm



Significantly mutated subnetworks and complexes
16 subnetworks containing 147 genes

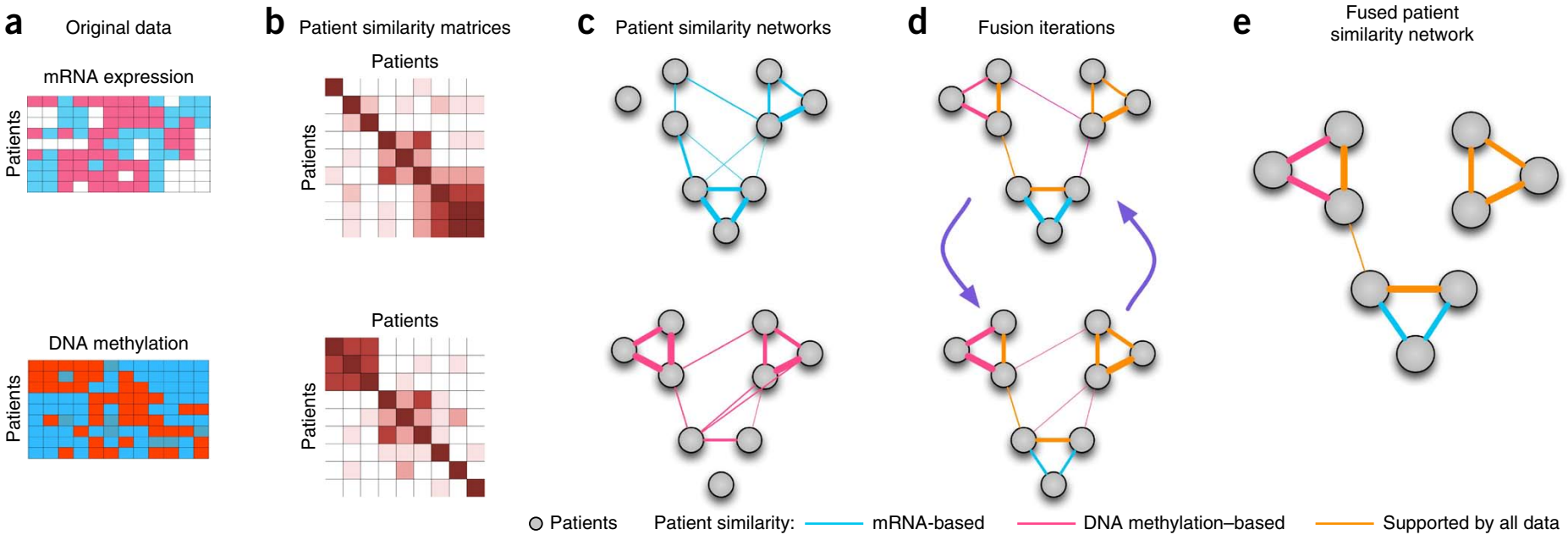


Computational concepts

1. Graph diffusion methods
2. Subnetwork identification

Ye et al, 2014, Science, Leiserson et al., Nature Genetics 2015

Graph-based data integration



Computational concepts

1. Graph clustering
2. Clustering multiple graphs

From Wang et al Nature Methods 2013

Plan for next few lectures

- Sep 11th , 13th
  - Background into graph theory, probability theory and molecular networks
  - Readings:
    - L. Hunter. Life and Its Molecules: A Brief Introduction. AI Magazine 25(1):9-22, 2004.
    - Winterbach et al., Topology of molecular interaction networks. BMC Systems Biology, 2013
      - Section on Network Biology
- Sep 15th
  - Probabilistic graphical models for molecular networks

Learning goals of this class

- Gain a broad overview of the application areas and computational solutions in Network biology
- Gain a deeper understanding one or two areas introduced
- Apply the computational concepts to similar problems in biology and complex systems
- Understand and critique scientific articles
- Enable self learning and deeper study of related topics

Goals for today

- Administrivia
- Course topics
- **Short survey of interests/background**